Final presentation and demo

Visual Analytics for Plant Pangenomes (VAPP)

Astrid van den Brandt | Eindhoven University of Technology
Anna Vilanova | Eindhoven University of Technology
Sandra Smit | Wageningen University & Research

5 October 2022

Agenda

1. **PanVA**: variant analysis within pangenomes
   - Design and process
   - Next steps
2. Demo on Pectobacterium
3. Questions

Problem statement (1)

Current focus

Exploring **sequence variation underlying traits**

1. within a target gene
2. across multiple accessions/strains
3. in the context of multiple references
4. hierarchical relations
5. phenotypes

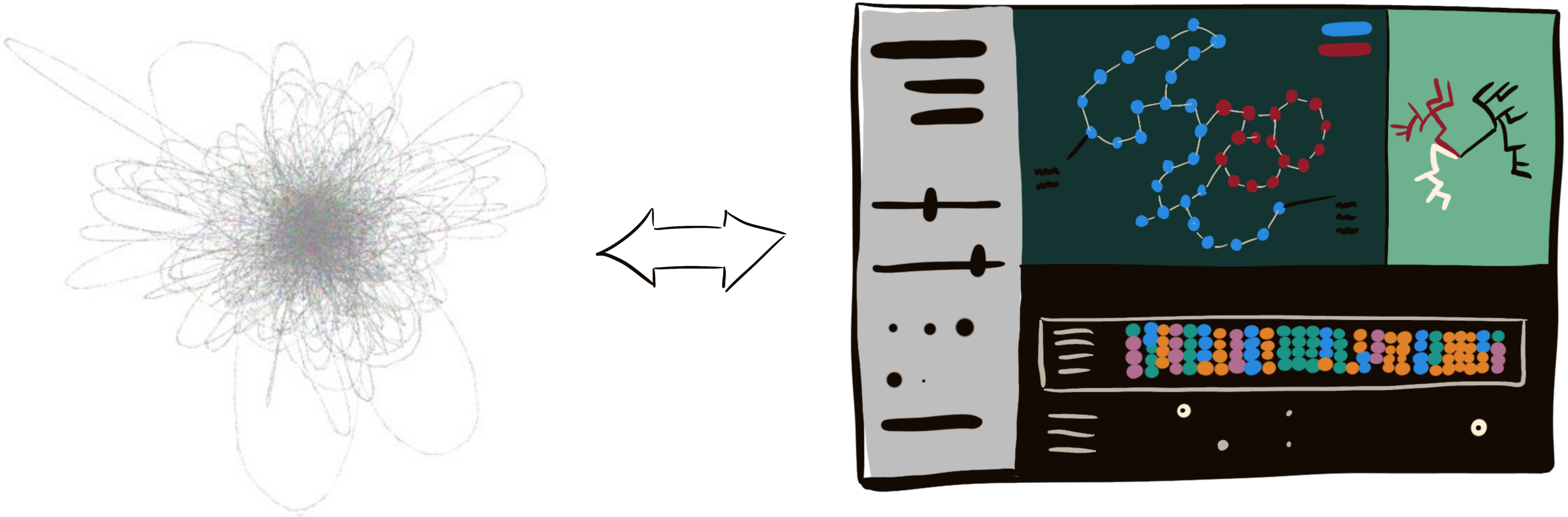
Pangenomes !!

Problem statement (2)

What is the problem?

1. Interpreting the graph in the biological context

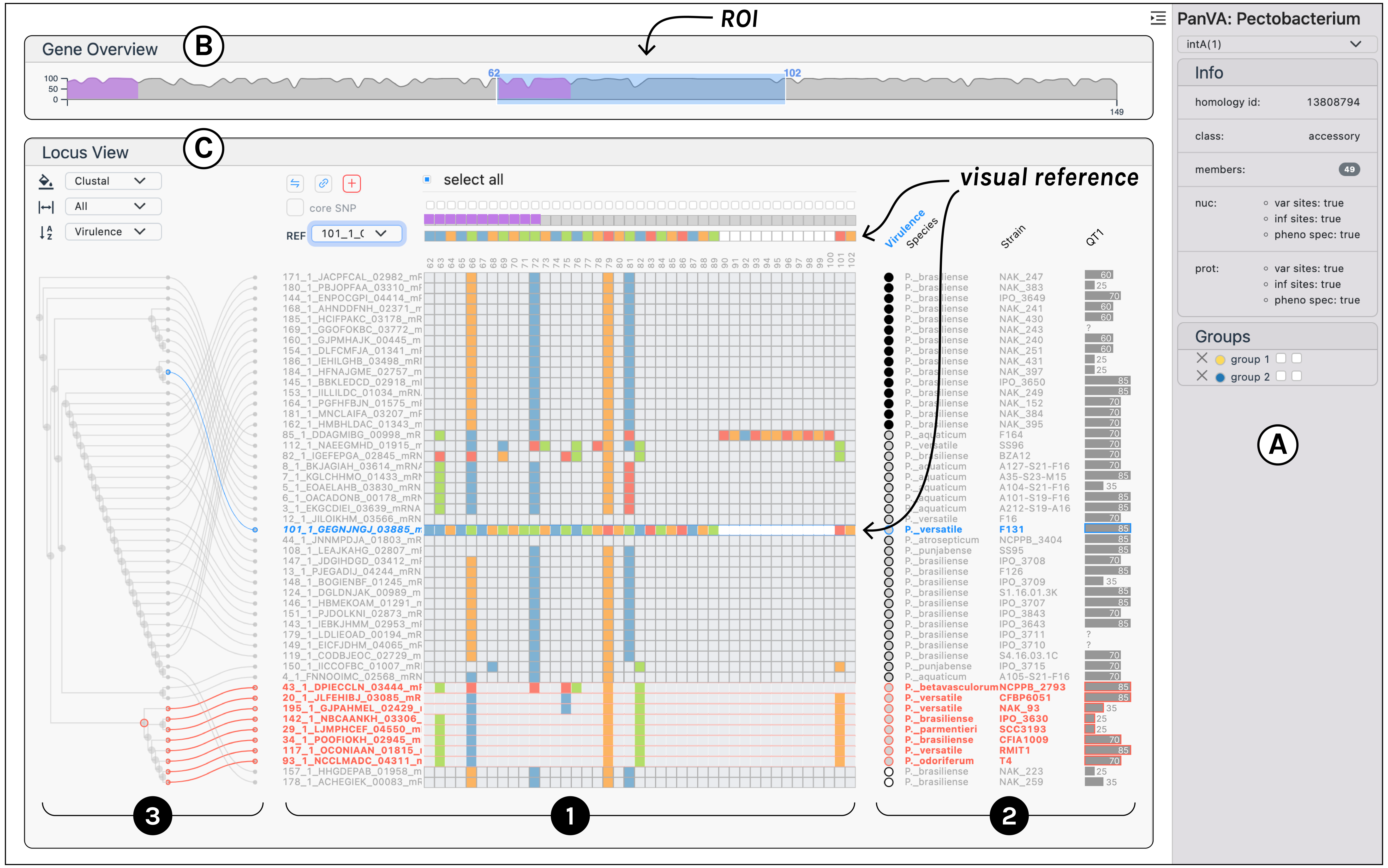
2. Current tools not sufficient for exploratory analysis
   - Only high level: focussed on displaying the graph
   - For specific organisms (e.g. microbes only)
   - Not possible to inspect metadata (and trees)



Translating pangenome information into meaningful views for genome scientists

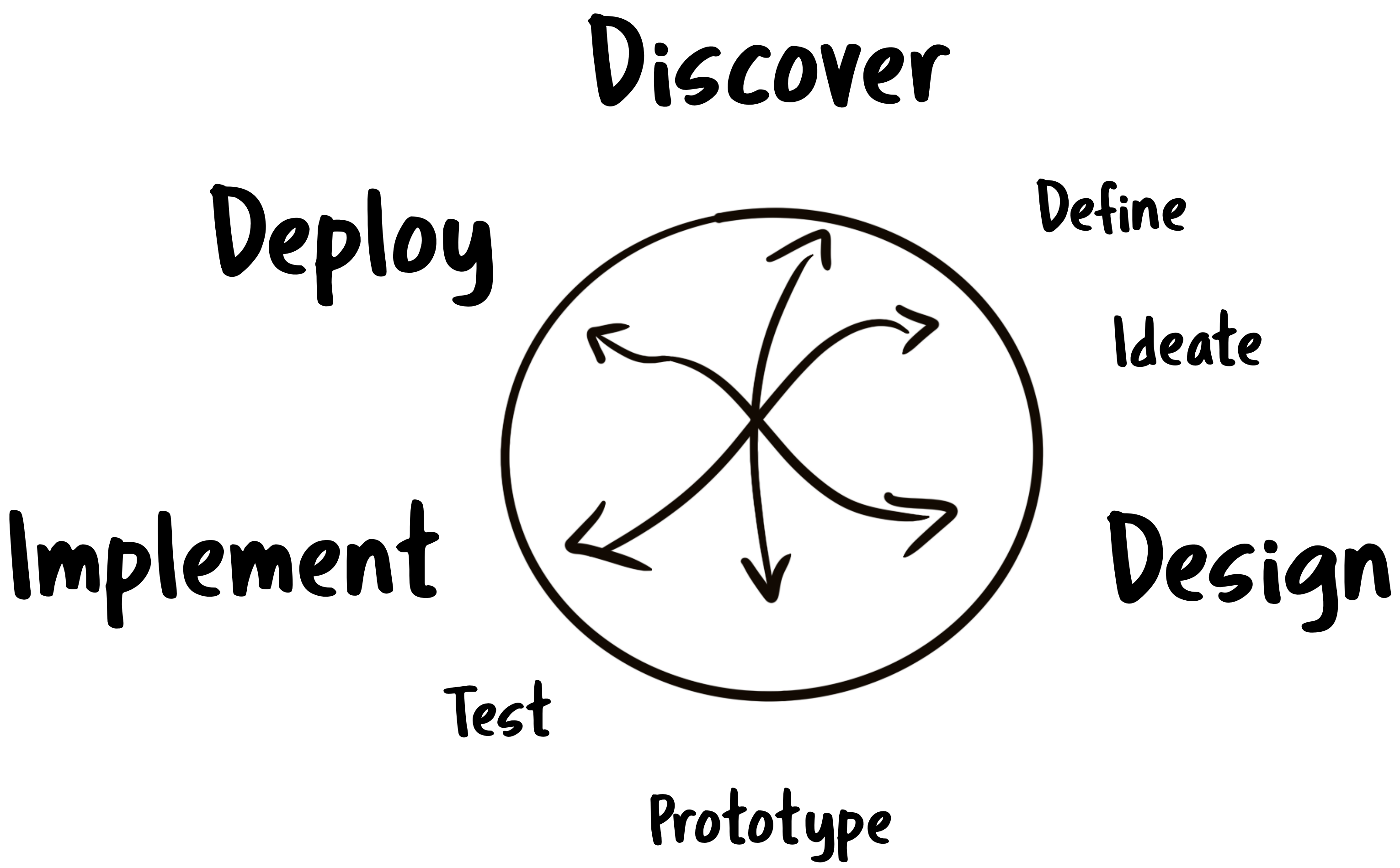
PanVA

Variant Analysis in Pangenomes



Tailored visual encodings and interactions (select, sort, group) to explore relations

Iterative UCD process

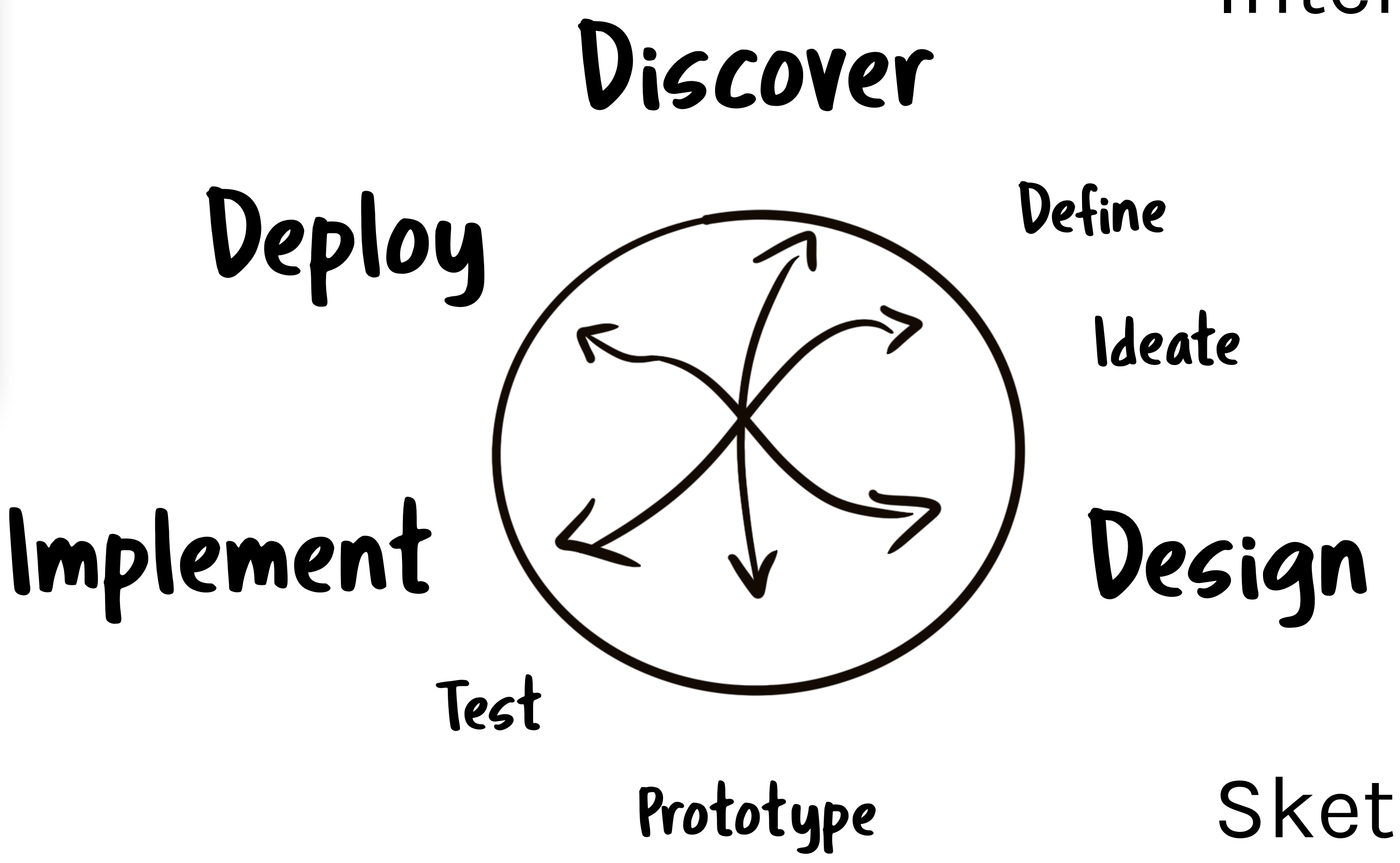


UCD = User Centered Design

Iterative UCD process

Prototype v1


Gene: AT1G02820
Visual Reference: Altai-5
Tree Type: k-mer distance
☑ Show branch length
Size Brush: 100 positions


Discover

Define

Ideate

Design

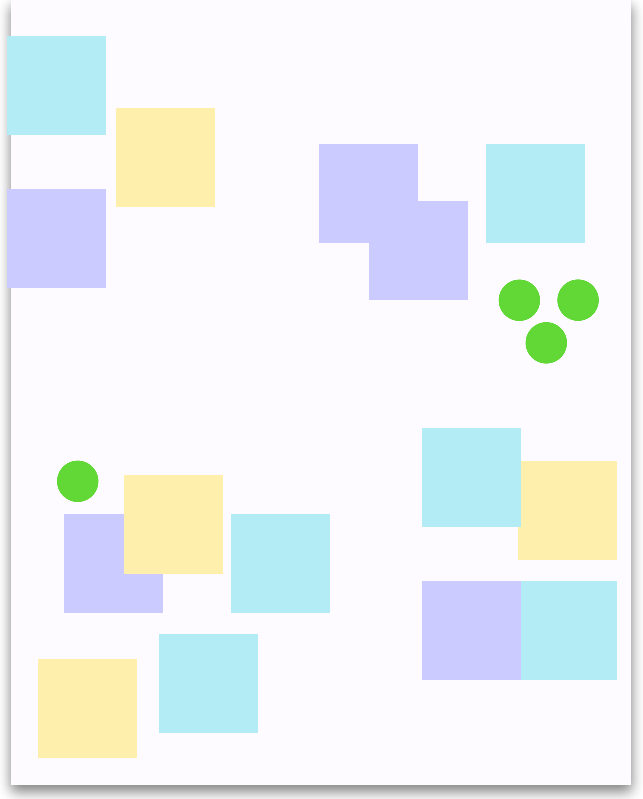
Prototype

Test

Implement

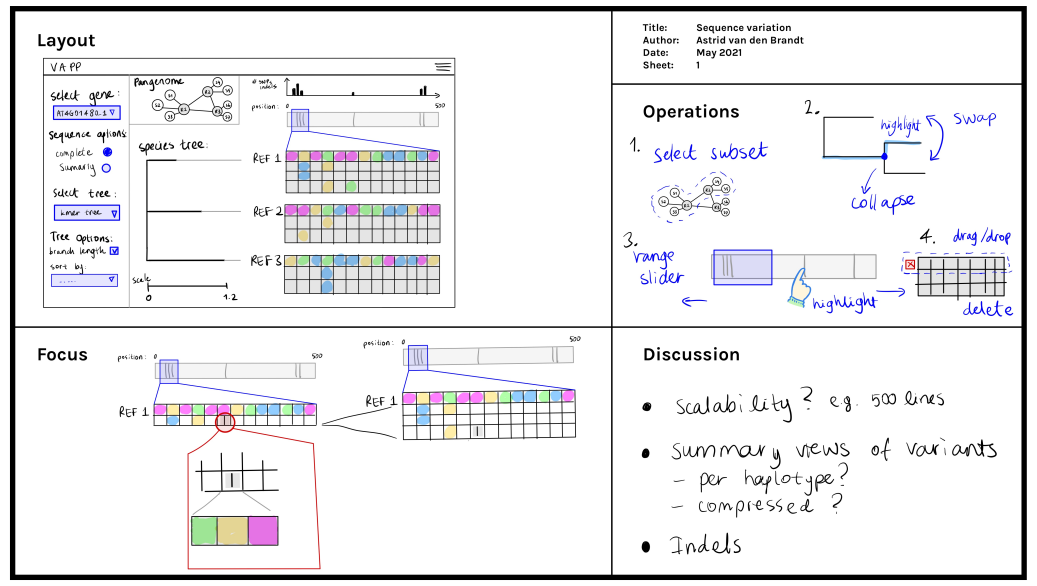
Deploy

Interviews



Creativity workshop

Sketches and mockups


Title: Sequence variation
Author: Astrid van den Brandt
Date: May 2021
Sheet: 1

Discussion

- Scalability? e.g. 500 lines
- summary views of variants
  - per haplotype?
  - compressed?
- Indels

UCD = User Centered Design

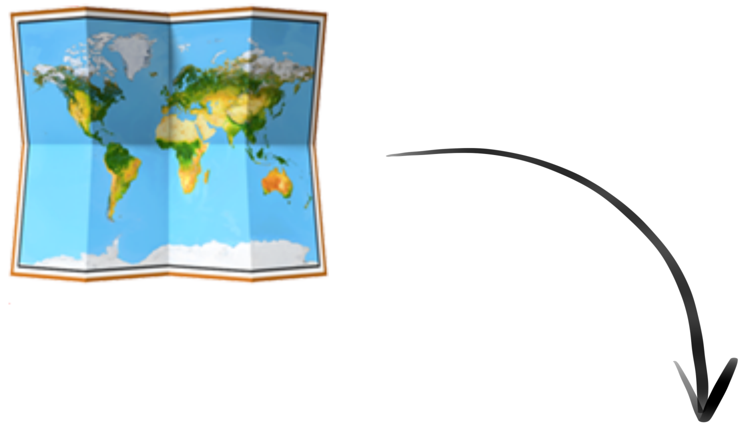
User Goals and Tasks (1)

G1. Freely explore genotype-phenotype relations

G2. Marker discovery

→ Exploration

→ Comparison



Gene / ROI

User Goals and Tasks (2)

🗺️

T1. Browse variation/conservation in ROI

🔬

T2. Examine sequences and annotations

T3. Explore phenotypes

T4. Explore (evolutionary) relations

⚖️

T5. Compare patterns

T6. Connect relationships by groups

T7. Summarise groups

PanVA: overview

A. Menu controls

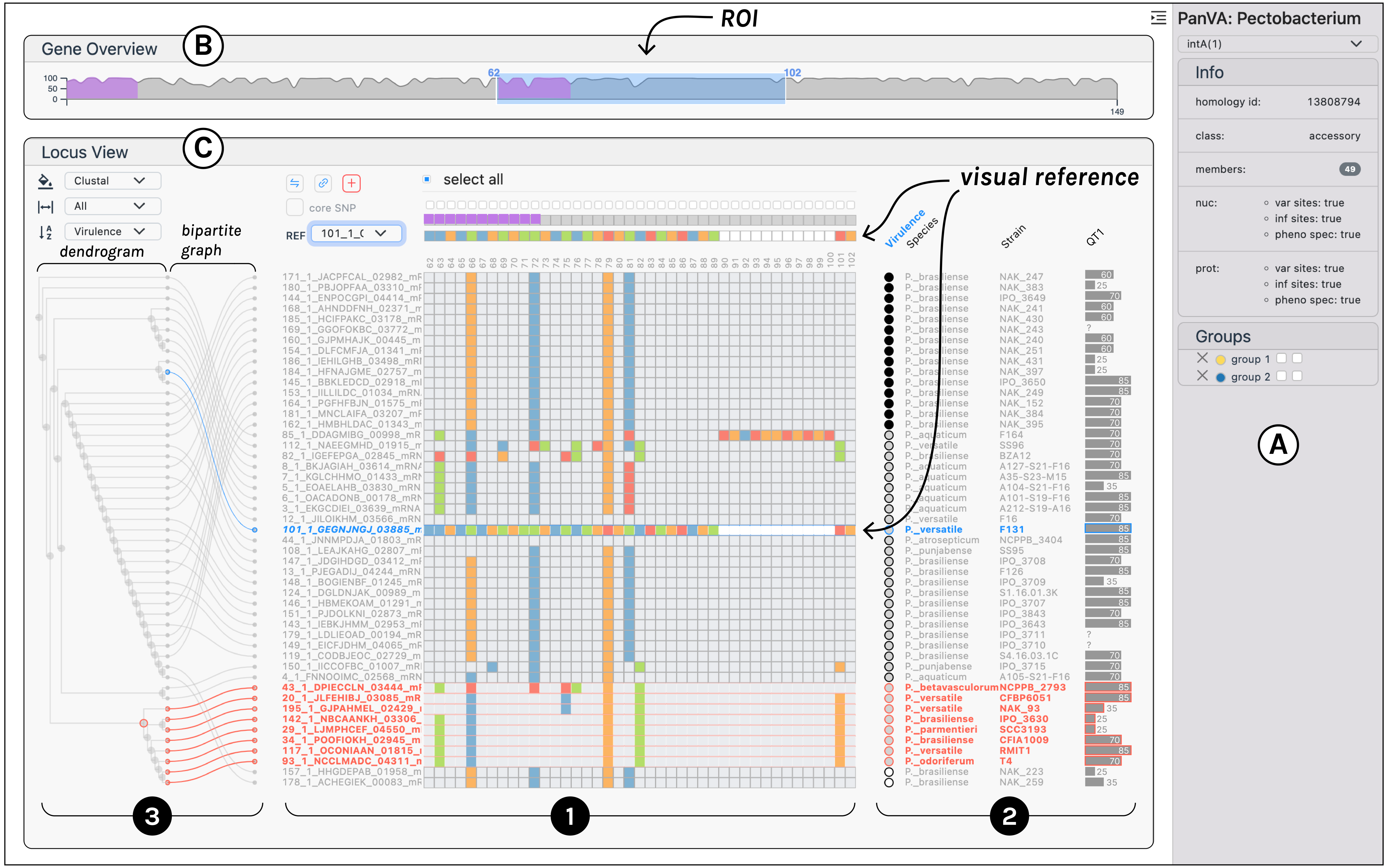
B. Gene Overview

C. Locus View

1. Sequences

2. Phenotypes

3. Relations



PanVA: Pectobacterium

intA(1)

Info

homology id: 13808794

class: accessory

members: 49

nuc:
- var sites: true
- inf sites: true
- pheno spec: true

prot:
- var sites: true
- inf sites: true
- pheno spec: true

Groups

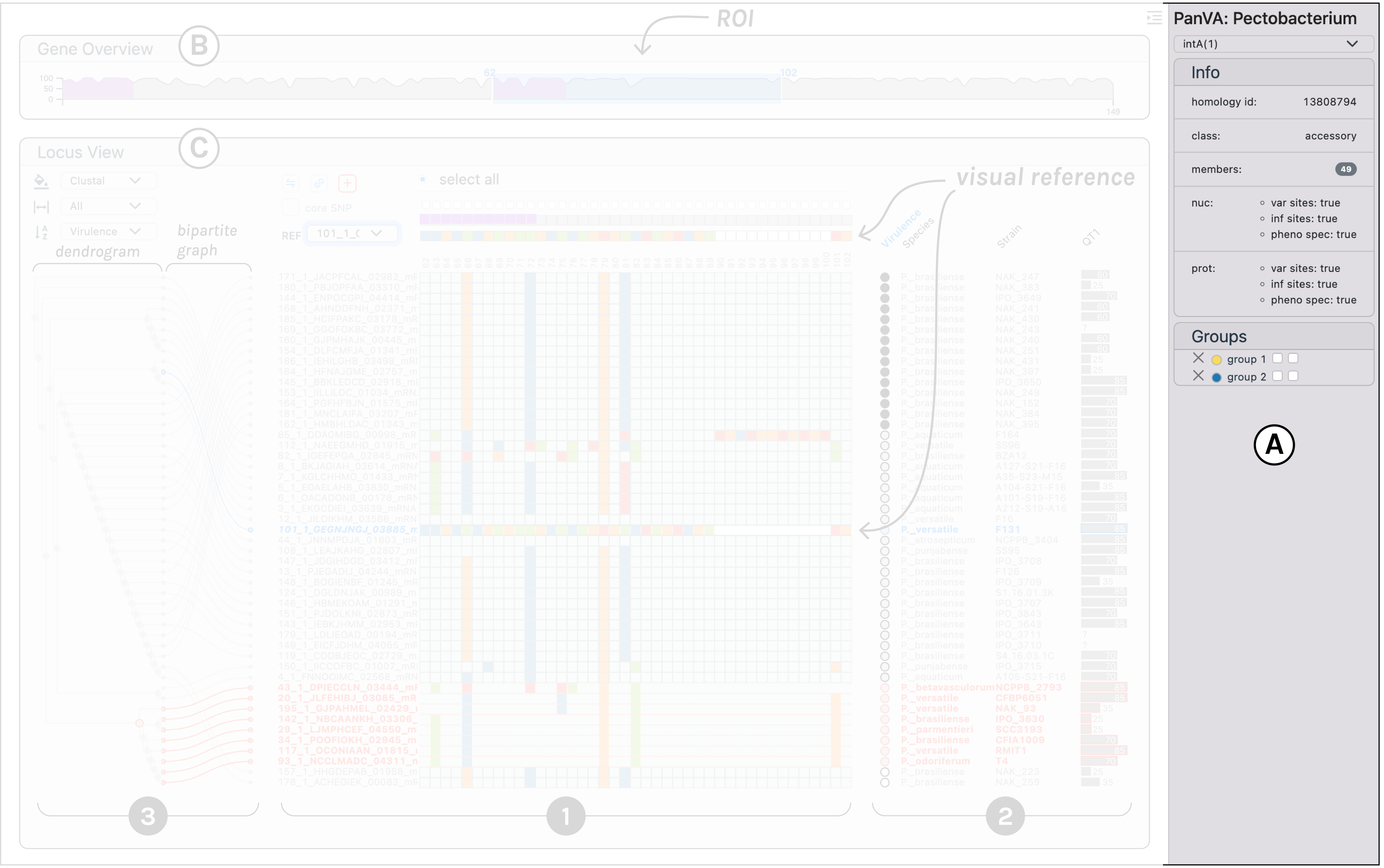
PanVA: overview

A. Menu controls

B. Gene Overview

C. Locus View

1. Sequences
2. Phenotypes
3. Relations



PanVA: Pectobacterium

intA(1)

homology id: 13808794

class: accessory

members: 49

nuc:
- var sites: true
- inf sites: true
- pheno spec: true

prot:
- var sites: true
- inf sites: true
- pheno spec: true

☐ ☐ group 1

☐ ☐ group 2

PanVA: overview

A. Menu controls

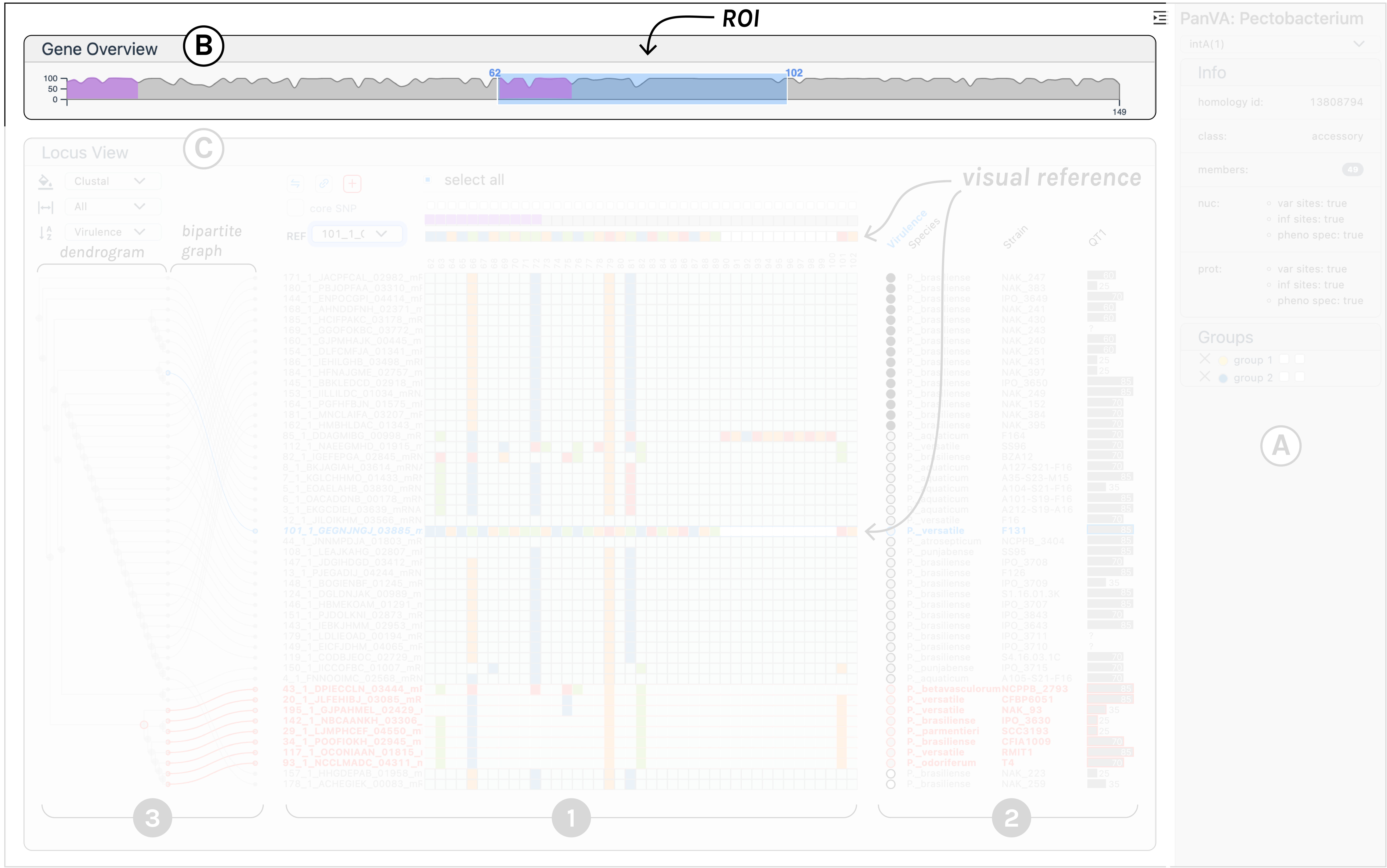
B. Gene Overview

C. Locus View

  1. Sequences

  2. Phenotypes

  3. Relations



PanVA: overview

A. Menu controls

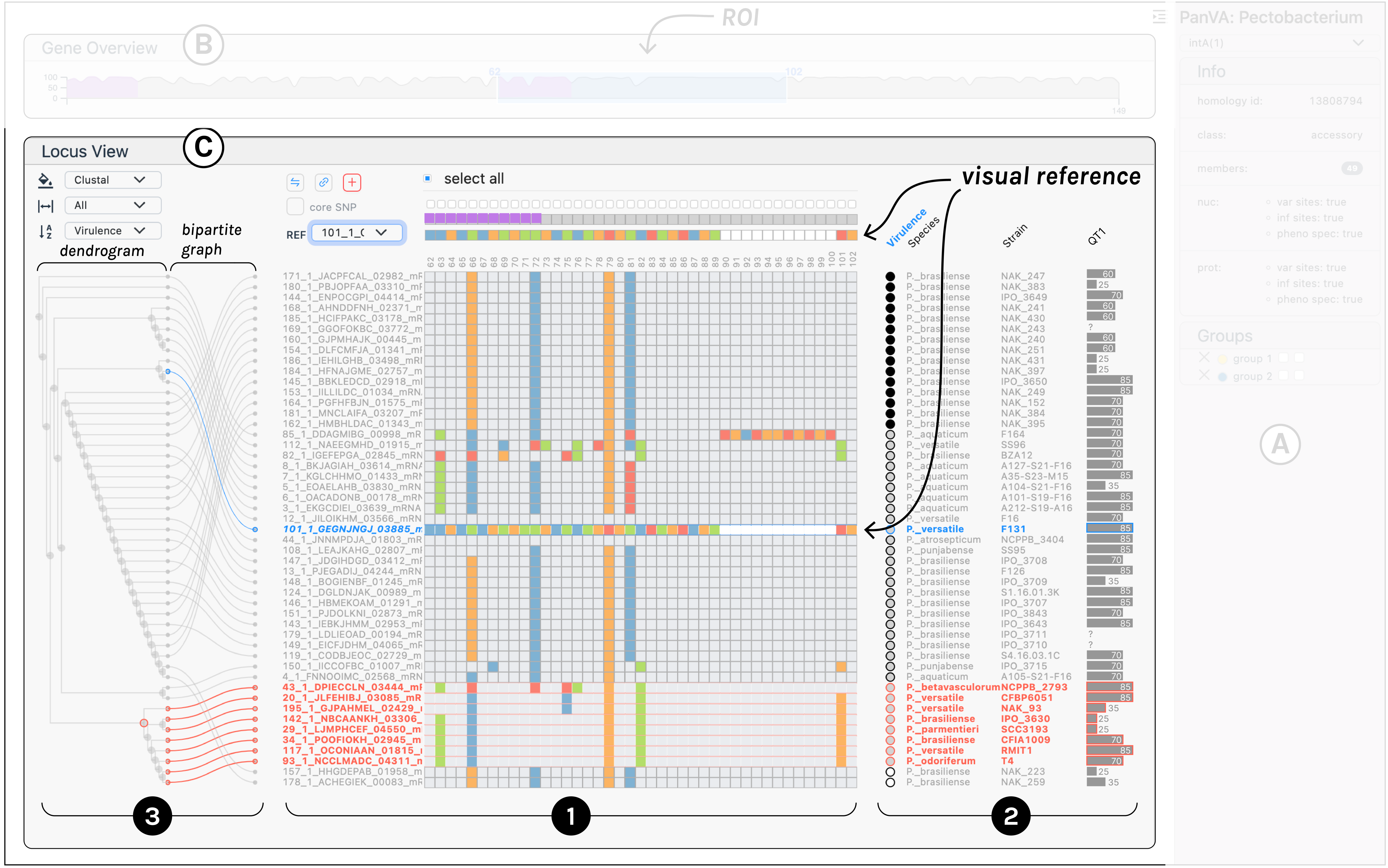
B. Gene Overview

C. Locus View

1. Sequences

2. Phenotypes

3. Relations



	Species	Strain	QT1
171_1_JACPFCAL_02982_mR	P._brasiliense	NAK_247	60
180_1_PBJOPFAA_03310_mR	P._brasiliense	NAK_383	25
144_1_ENPOCGPI_04414_mR	P._brasiliense	IPO_3649	70
168_1_AHNDDFNH_02371_m	P._brasiliense	NAK_241	60
185_1_HCIFPAKC_03178_mR	P._brasiliense	NAK_430	60
169_1_GGOFOKBC_03772_m	P._brasiliense	NAK_243	?
160_1_GJPMHAJK_00445_m	P._brasiliense	NAK_240	60
154_1_DLFCMFJA_01341_mR	P._brasiliense	NAK_251	60
186_1_IEHILGHB_03498_mR	P._brasiliense	NAK_431	25
184_1_HFNAJGME_02757_m	P._brasiliense	NAK_397	25
145_1_BBKLEDCD_02918_m	P._brasiliense	IPO_3650	85
153_1_IILLILDC_01034_mRN	P._brasiliense	NAK_249	85
164_1_PGFHFBJN_01575_mR	P._brasiliense	NAK_152	70
181_1_MNCLAIFA_03207_mR	P._brasiliense	NAK_384	70
162_1_HMBHLDAC_01343_m	P._brasiliense	NAK_395	70
85_1_DDAGMIBG_00998_mR	P._aquaticum	F164	70
112_1_NAEEGMHD_01915_m	P._versatile	SS96	70
82_1_IGEFEPGA_02845_mRN	P._brasiliense	BZA12	70
8_1_BKJAGIAH_03614_mRN	P._aquaticum	A127-S21-F16	70
7_1_KGLCHHMO_01433_mRN	P._aquaticum	A35-S23-M15	85
5_1_EOAELAHB_03830_mRN	P._aquaticum	A104-S21-F16	35
6_1_OACADONB_00178_mRN	P._aquaticum	A101-S19-F16	85
3_1_EKGCDIEI_03639_mRNA	P._aquaticum	A212-S19-A16	85
12_1_JILOIKHM_03566_mRN	P._versatile	F16	70
101_1_GEGNJNGJ_03885_m	**P._versatile**	**F131**	85
44_1_JNNMPDJA_01803_mR	P._atrosepticum	NCPPB_3404	85
108_1_LEAJKAHG_02807_mR	P._punjabense	SS95	85
147_1_JDGIHDGD_03412_mR	P._brasiliense	IPO_3708	70
13_1_PJEGADIJ_04244_mRN	P._brasiliense	F126	85
148_1_BOGIENBF_01245_mR	P._brasiliense	IPO_3709	35
124_1_DGLDNJAK_00989_m	P._brasiliense	S1.16.01.3K	85
146_1_HBMEKOAM_01291_m	P._brasiliense	IPO_3707	85
151_1_PJDOLKNI_02873_mR	P._brasiliense	IPO_3843	70
143_1_IEBKJHMM_02953_mR	P._brasiliense	IPO_3643	85
179_1_LDLIEOAD_00194_mR	P._brasiliense	IPO_3711	?
149_1_EICFJDHM_04065_mR	P._brasiliense	IPO_3710	?
119_1_CODBJEOC_02729_m	P._brasiliense	S4.16.03.1C	70
150_1_IICCOFBC_01007_mR	P._punjabense	IPO_3715	70
4_1_FNNOOIMC_02568_mRN	P._aquaticum	A105-S21-F16	70
43_1_DPIECCLN_03444_mR	**P._betavasculorum**	**NCPPB_2793**	85
20_1_JLFEHIBJ_03085_mR	**P._versatile**	**CFBP6051**	85
195_1_GJPAHMEL_02429_	**P._versatile**	**NAK_93**	35
142_1_NBCAANKH_03306_	**P._brasiliense**	**IPO_3630**	25
29_1_LJMPHCEF_04550_m	**P._parmentieri**	**SCC3193**	25
34_1_POOFIOKH_02945_m	**P._brasiliense**	**CFIA1009**	70
117_1_OCONIAAN_01815_	**P._versatile**	**RMIT1**	85
93_1_NCCLMADC_04311_n	**P._odoriferum**	**T4**	70
157_1_HHGDEPAB_01958_m	P._brasiliense	NAK_223	25
178_1_ACHEGIEK_00083_mR	P._brasiliense	NAK_259	35

PanVA: overview

A. Menu controls

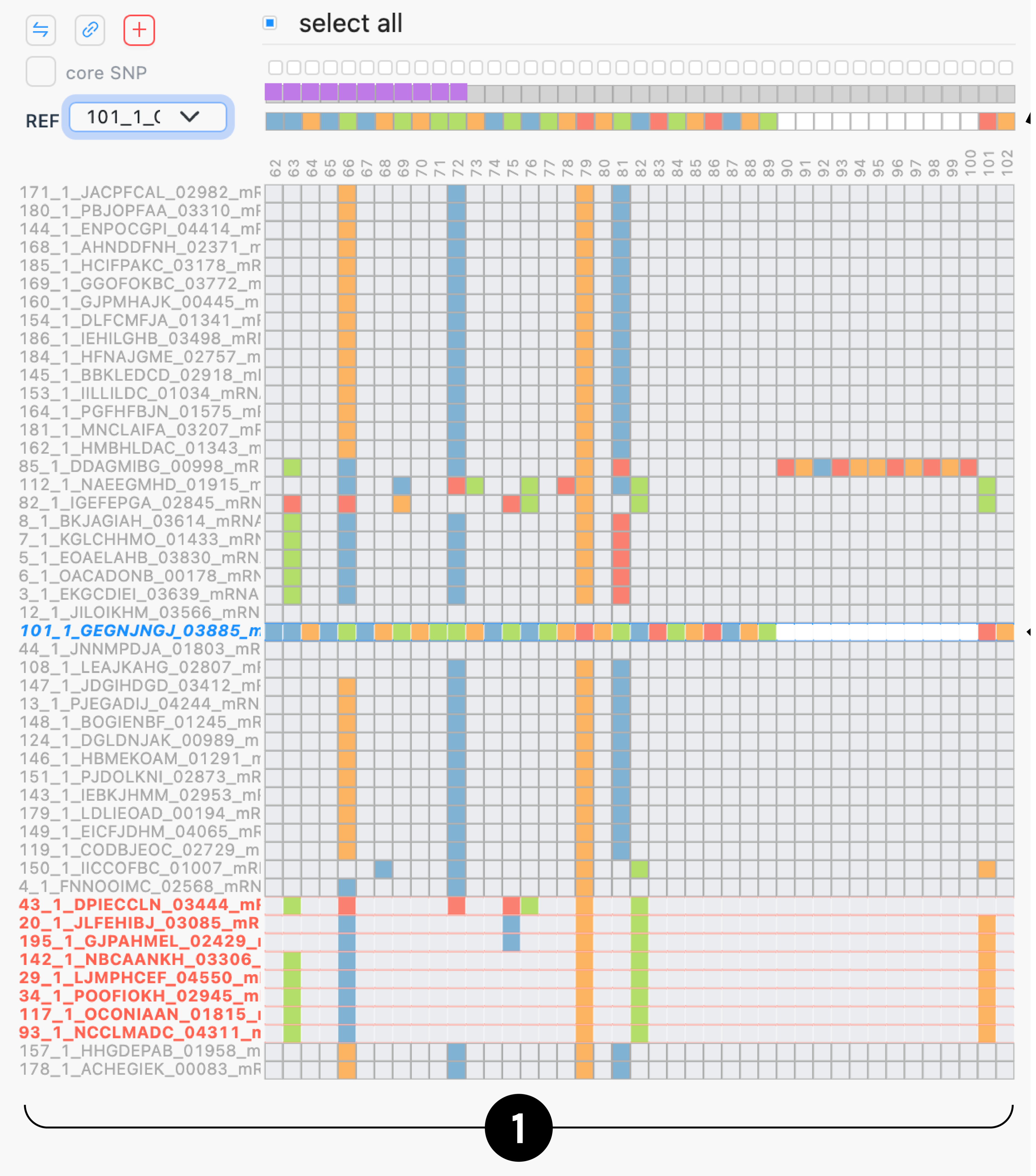
B. Gene Overview

C. Locus View

  1. Sequences

  2. Phenotypes

  3. Relations



PanVA: overview

A. Menu controls

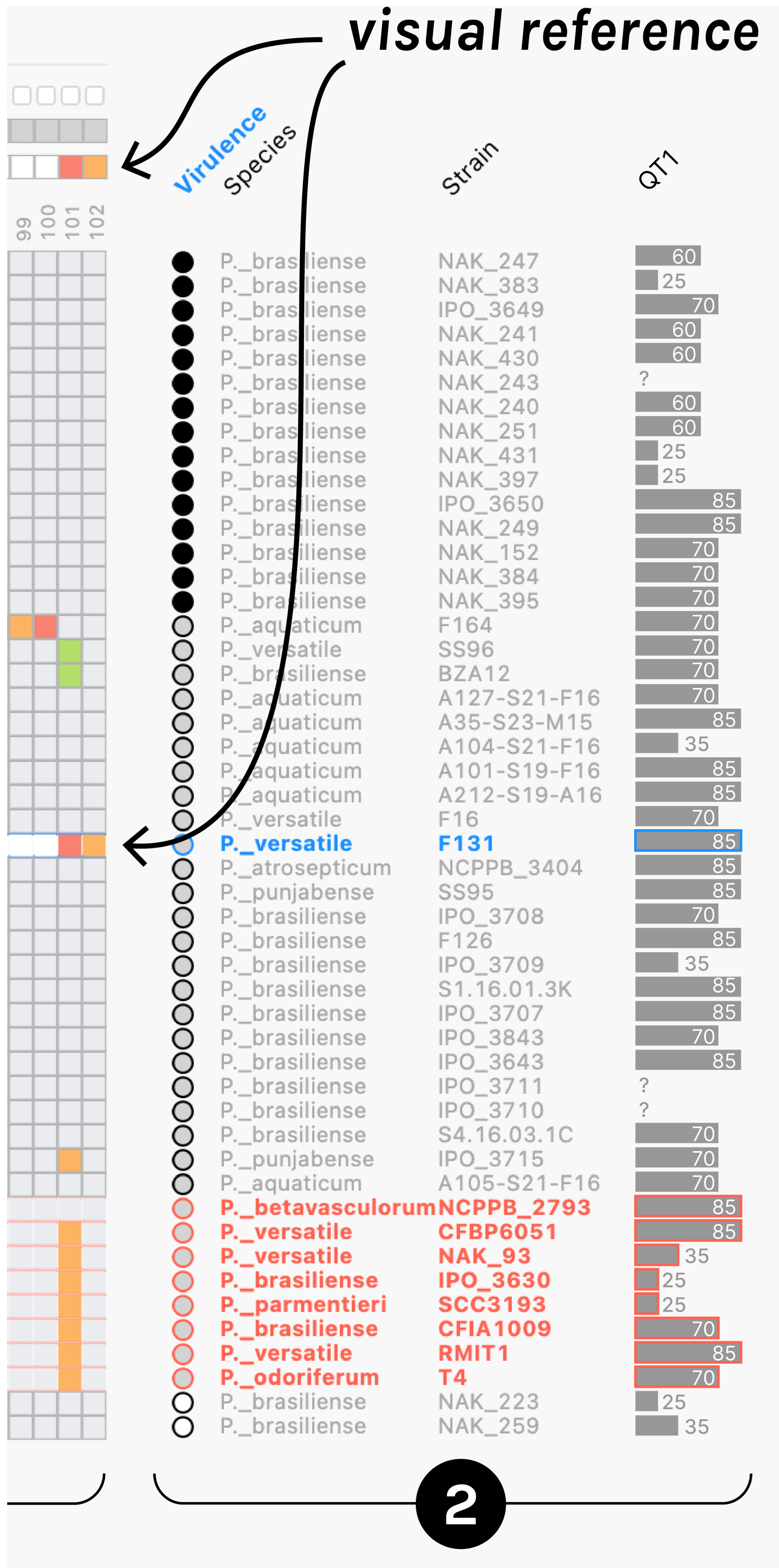
B. Gene Overview

C. Locus View

1. Sequences

2. Phenotypes

3. Relations



Virulence	Species	Strain	QT1
●	P._brasiliense	NAK_247	60
●	P._brasiliense	NAK_383	25
●	P._brasiliense	IPO_3649	70
●	P._brasiliense	NAK_241	60
●	P._brasiliense	NAK_430	60
●	P._brasiliense	NAK_243	?
●	P._brasiliense	NAK_240	60
●	P._brasiliense	NAK_251	60
●	P._brasiliense	NAK_431	25
●	P._brasiliense	NAK_397	25
●	P._brasiliense	IPO_3650	85
●	P._brasiliense	NAK_249	85
●	P._brasiliense	NAK_152	70
●	P._brasiliense	NAK_384	70
●	P._brasiliense	NAK_395	70
○	P._aquaticum	F164	70
○	P._versatile	SS96	70
○	P._brasiliense	BZA12	70
○	P._aquaticum	A127-S21-F16	70
○	P._aquaticum	A35-S23-M15	85
○	P._aquaticum	A104-S21-F16	35
○	P._aquaticum	A101-S19-F16	85
○	P._aquaticum	A212-S19-A16	85
○	P._versatile	F16	70
○	**P._versatile**	**F131**	85
○	P._atrosepticum	NCPPB_3404	85
○	P._punjabense	SS95	85
○	P._brasiliense	IPO_3708	70
○	P._brasiliense	F126	85
○	P._brasiliense	IPO_3709	35
○	P._brasiliense	S1.16.01.3K	85
○	P._brasiliense	IPO_3707	85
○	P._brasiliense	IPO_3843	70
○	P._brasiliense	IPO_3643	85
○	P._brasiliense	IPO_3711	?
○	P._brasiliense	IPO_3710	?
○	P._brasiliense	S4.16.03.1C	70
○	P._punjabense	IPO_3715	70
○	P._aquaticum	A105-S21-F16	70
○	**P._betavasculorum**	**NCPPB_2793**	85
○	**P._versatile**	**CFBP6051**	85
○	**P._versatile**	**NAK_93**	35
○	**P._brasiliense**	**IPO_3630**	25
○	**P._parmentieri**	**SCC3193**	25
○	**P._brasiliense**	**CFIA1009**	70
○	**P._versatile**	**RMIT1**	85
○	**P._odoriferum**	**T4**	70
○	P._brasiliense	NAK_223	25
○	P._brasiliense	NAK_259	35

PanVA: overview

A. Menu controls

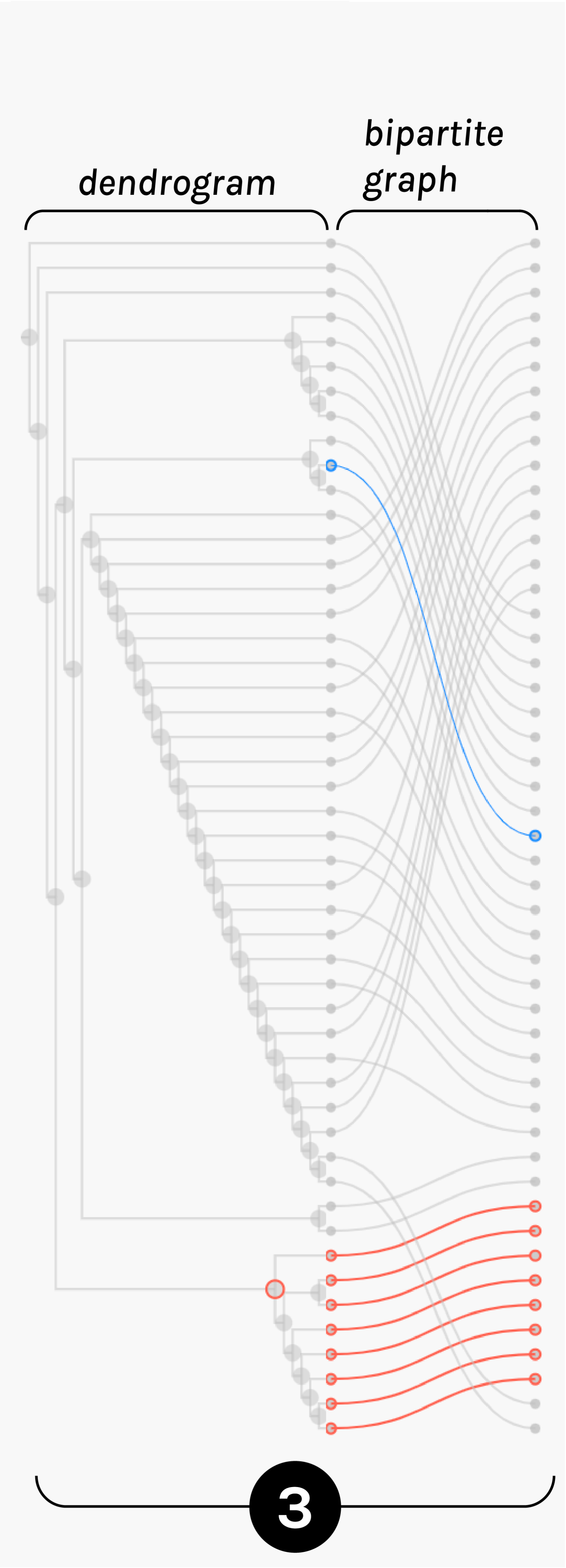
B. Gene Overview

C. Locus View

  1. Sequences

  2. Phenotypes

  3. Relations



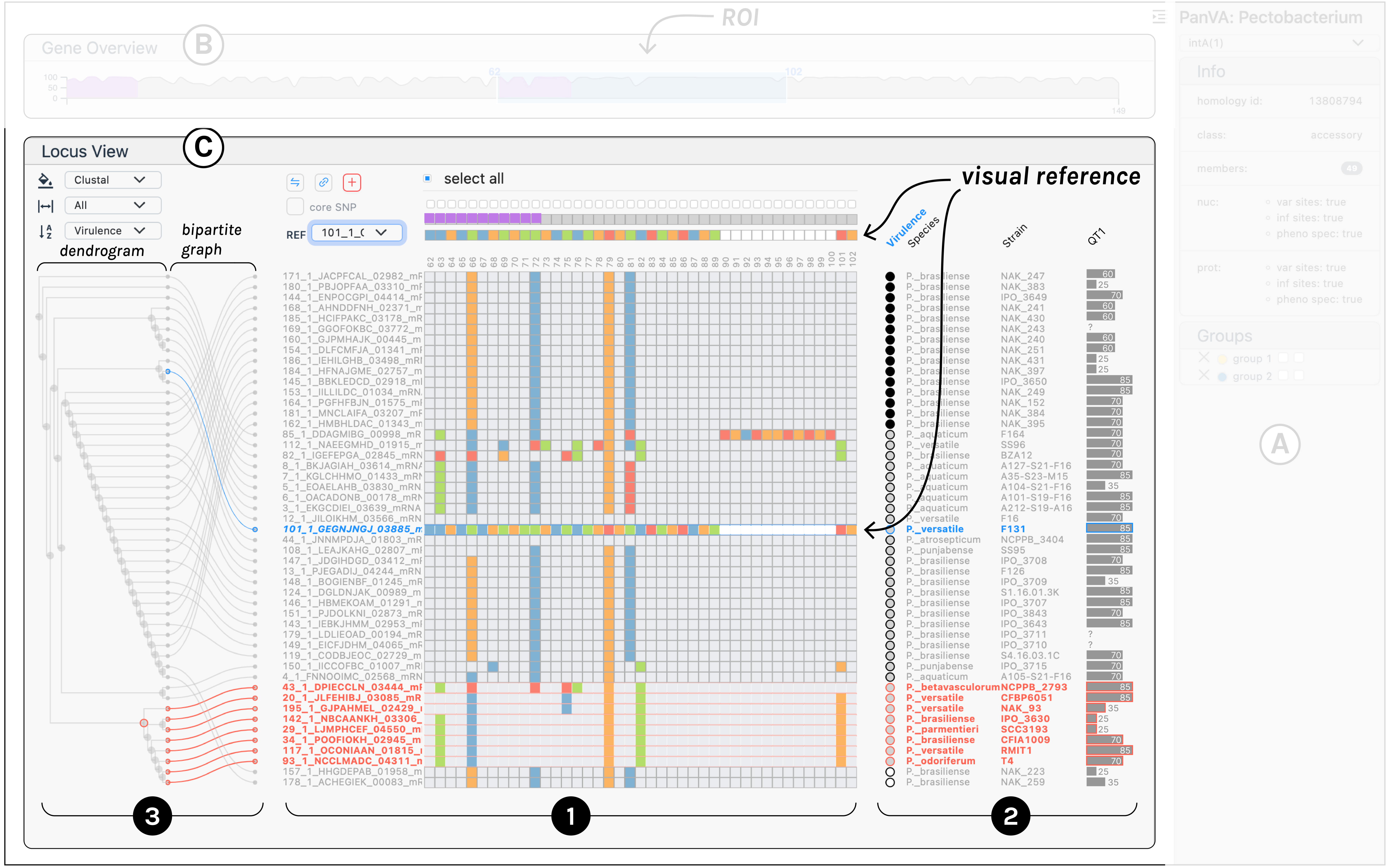
PanVA: overview

A. Menu controls

B. Gene Overview

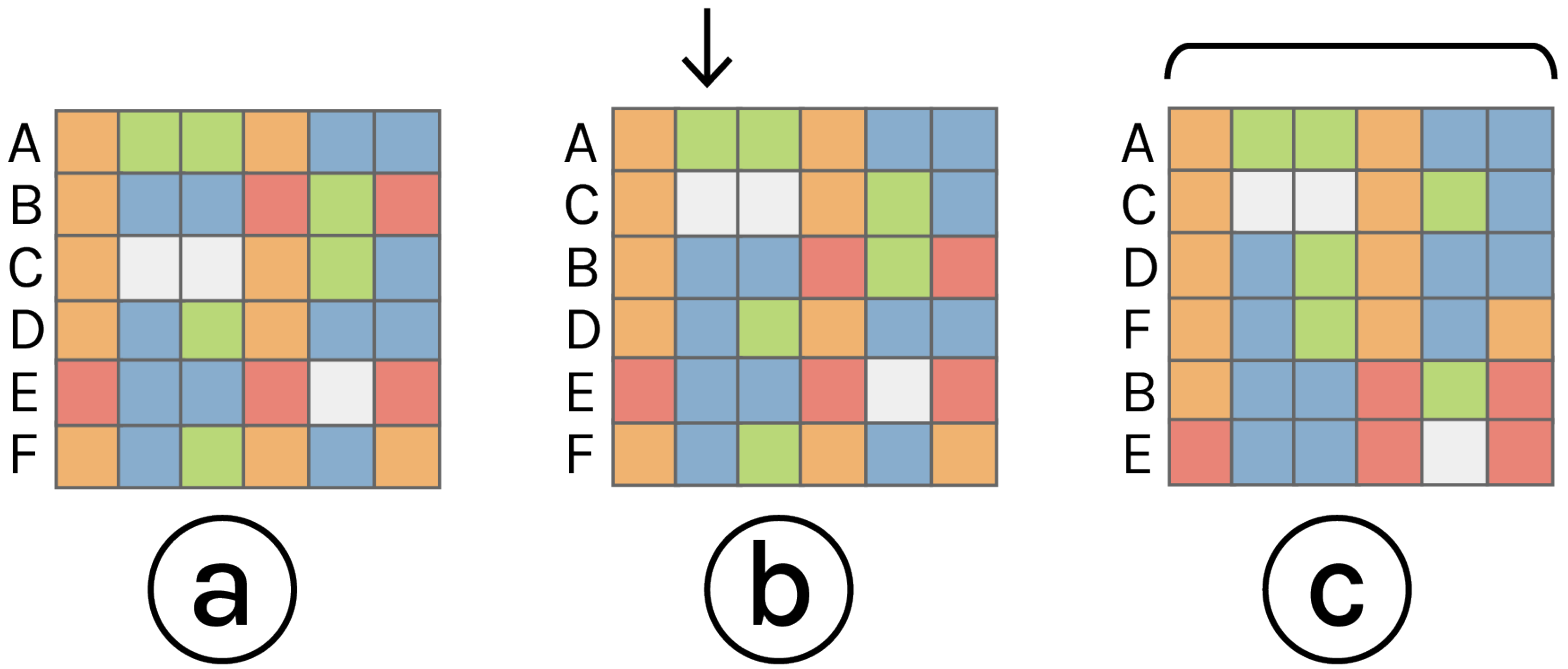
C. Locus View

1. Sequences
2. Phenotypes
3. Relations



PanVA: interactions (1)

Sorting



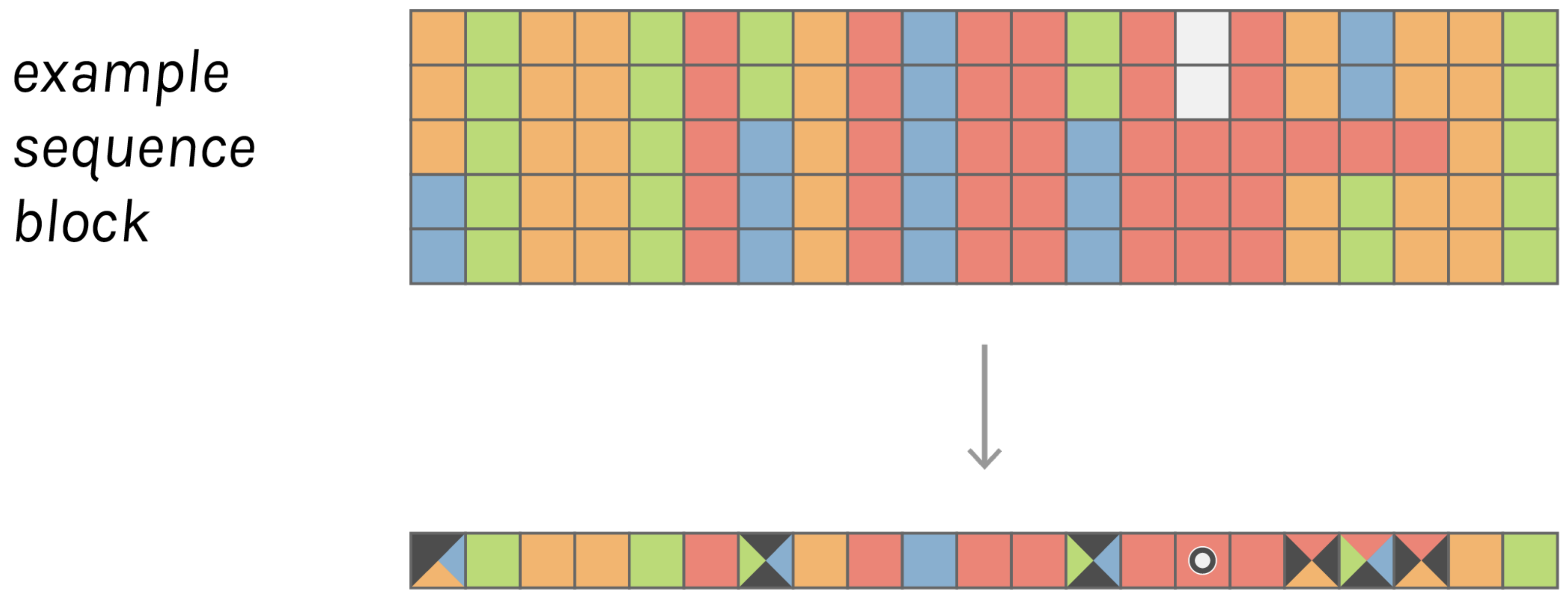
None

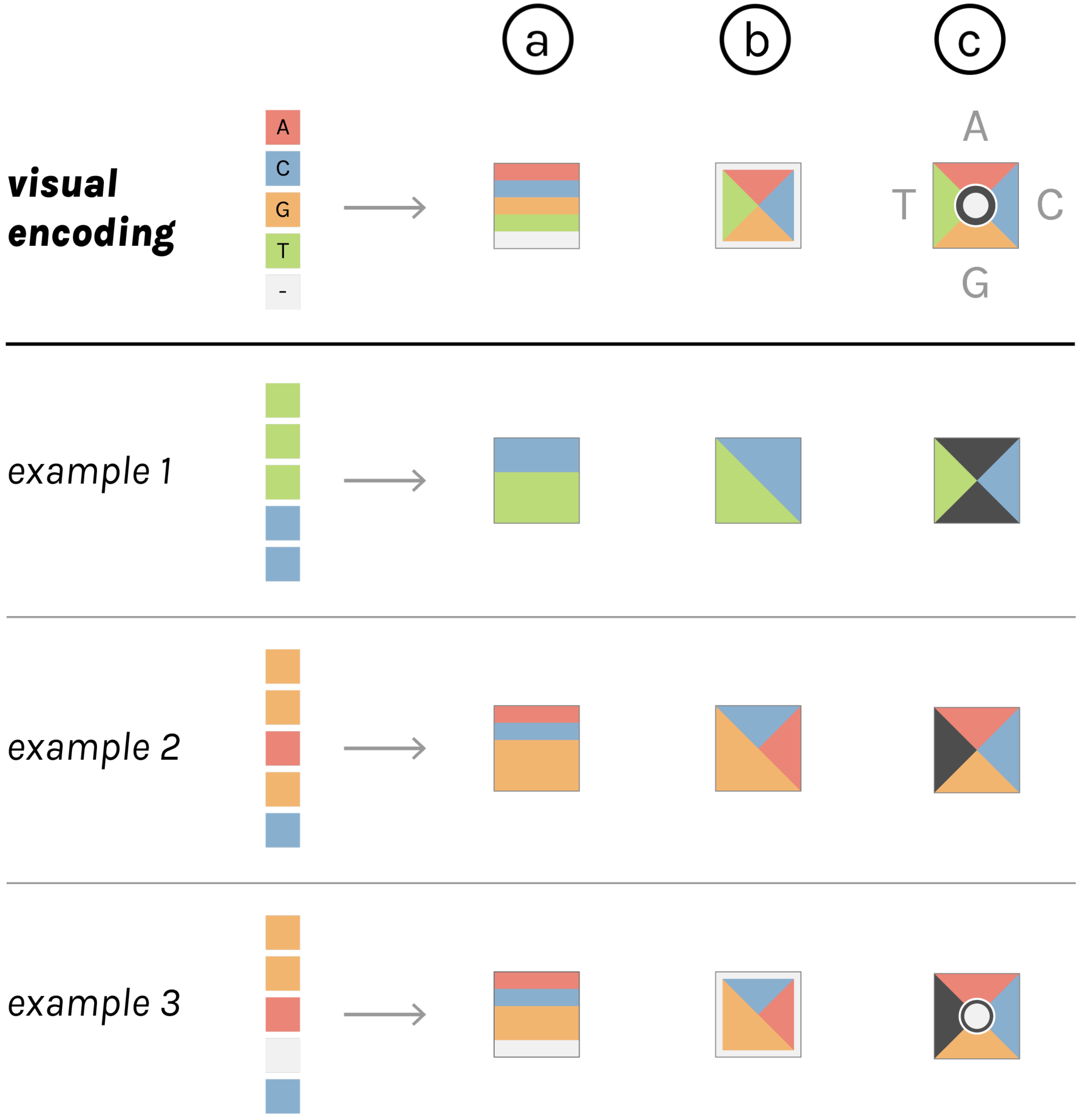
Single position

Multi position

PanVA: interactions (2)

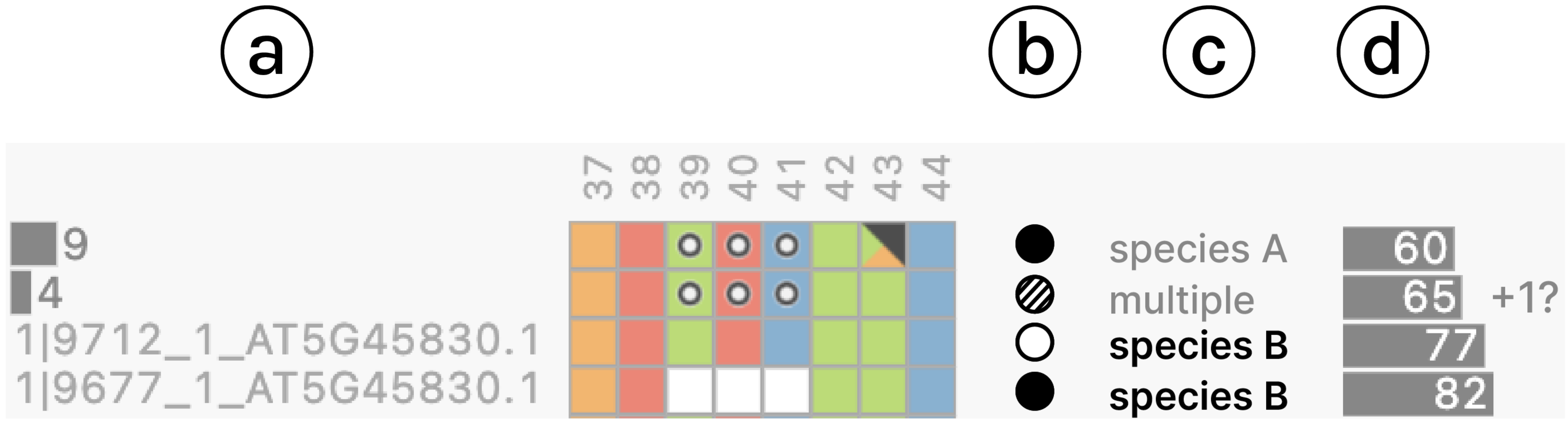
Grouping + aggregation





PanVA: interactions (2)

Grouping + aggregation



Labels **Sequences** **Phenotypes**

PanVA: implementation

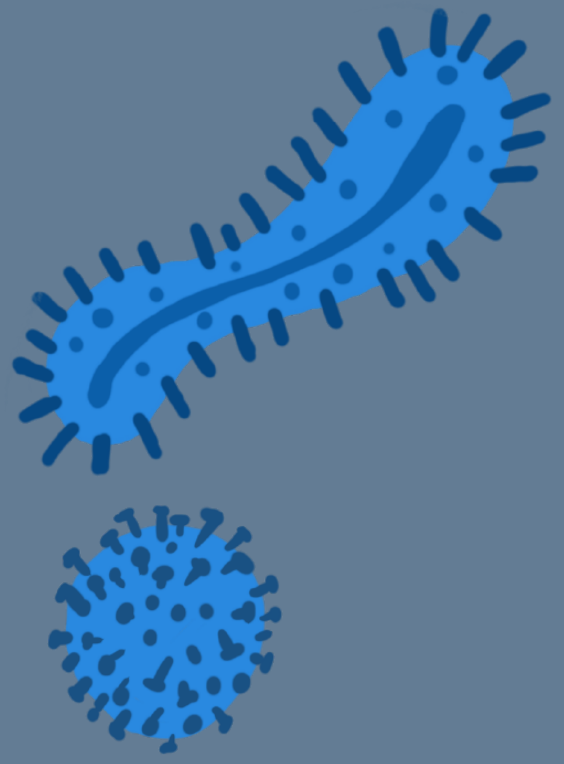
- Frontend: Vue + D3.js
- Backend: Python + Flask

- 2 demo instances: Arabidopsis & Pectobacterium

PanVA: evaluation

- Use cases
  - Pectobacterium barcode genes
  - Arabidopsis Flowering Time genes

- Qualitative feedback sessions
  - Interactions (grouping) - Integration
  - Glyph



Live demo

Pectobacterium barcode genes

PanVA: next steps

✏️ **Design**

- Usability aspects (e.g. groups from phenotypes)
- Visual scalability (e.g. zooming)
    - Scaling to larger regions (horizontally) and to more sequences (vertically)
- Large scale variation (synteny and PAV)

⚙️ **Performance and integration with PanTools (eTEC)**

- Code refactoring
- Computational scalability: performance and rendering
- Connection with PanTools DB and instances